Bridge concrete spalling recognition based on generative AI synthetic weather and YOLOv11

Rong-Lu Hong[1], KiWon Lee[1], SeoYoung Park[1], Seunghyeon Wang[1] and Ju-Hyung Kim[1]

[1] Department of Architectural Engineering, Hanyang University, Seoul, Republic of Korea
abraham0814@hanyang.ac.kr, fosys8651@hanmail.net, ondamco@gmail.com, wsh1019@hanyang.ac.kr , kcr97jhk@hanyang.ac.kr

Abstract –

Automatic concrete spalling detection on bridge surfaces using unmanned aerial vehicles (UAV) is essential for improving inspection efficiency and operational safety. However, instance segmentation models employed in UAV-based inspections are sensitive to environment-related variations in image appearance. In practice, inspection data are usually collected under limited conditions, which reduces model reliability when applied beyond standard operating scenarios. To address this limitation, weather condition modelling strategies were examined from a data construction perspective. Three datasets derived from the same UAV inspection imagery were used: a baseline, a traditionally augmented dataset based on global statistical transformations, and a generative-AI-based dataset representing diverse inspection environments. All datasets shared identical pixel-level annotations. A unified instance segmentation model, YOLOv11 seg, was trained under identical configurations and evaluated using cross-dataset testing to assess generalization under simulated weather variability. Results demonstrate that the model trained with generative-AI-enhanced data achieves stable segmentation performance in challenging inspection environments. Meanwhile, the segmentation accuracy on the original test set remained high, indicating that improved environmental adaptability was achieved without sacrificing performance under regular inspection conditions. These findings demonstrate that generative-AI-based environment synthesis effectively reduces training data bias and enhances model reliability without additional annotation, thereby providing a practical data construction strategy for UAV-based bridge inspection in real-world applications.

Keywords –

Bridge inspection; Concrete spalling; Generative AI; Synthetic weather data; YOLOv11 segmentation

1 Introduction

Bridges are critical components of transportation infrastructure, and their service conditions directly affect public safety and the stable operation of transportation systems [1]. Surface concrete spalling is a typical bridge deterioration and damage form [2,3]. Such damage compromises the concrete cover and accelerates reinforcement corrosion, thereby reducing structural load-bearing capacity and durability [3]. According to the 2025 National Bridge Inventory of the United States, approximately 620,000 highway bridges are currently in service, more than 260,000 of which exhibit structural deficiencies, with an estimated investment of approximately $467 billion USD required for comprehensive rehabilitation [4]. If surface damage such as concrete spalling is not identified and repaired at an early stage, it often progresses to structural deterioration, substantially increasing maintenance costs and operational risks [1–3].

In current bridge operation and maintenance systems, surface concrete spalling identification relies primarily on manual inspection methods [3,5]. Such inspections typically require close-range visual examinations using aerial work equipment and exhibit limitations in inspection efficiency, operational safety, and result consistency, with outcomes largely dependent on inspector experience and judgment [1,2,5–7]. With the development of unmanned aerial vehicle (UAV) platforms and computer vision technologies, automated bridge surface damage detection based on UAV imagery has become a feasible technical approach [1,6,7]. Combined with deep-learning-based instance segmentation models, fine-grained localization of concrete spalling regions can be achieved under non-contact conditions, thereby improving inspection efficiency and reducing inspection risk [6,8]. However, the effectiveness of these methods is highly sensitive to training data quality and diversity [9]. In actual engineering environments, variations in weather and lighting conditions significantly affect the appearance of bridge imagery, limiting the model applicability under complex conditions [10,11].

To alleviate the shortage of data under complex environmental conditions, prior studies have expanded original imagery using conventional image augmentation techniques, including global statistical transformations such as brightness and contrast adjustment and noise perturbation [10,11]. Although these methods can increase the appearance diversity of training samples to a certain extent, their image level operations make it difficult to represent the local texture variations induced by weather conditions and the spatial interactions between such variations and structural surfaces [11,12]. In recent years, generative methods have gradually been introduced for the synthesis of complex scene samples. Style transfer networks such as CycleGAN have been applied to simulate road or structural scenes under different weather conditions [13,14]. However, the domain transformation process may introduce geometric distortions, making it difficult to preserve the spatial consistency of damage boundaries. Diffusion models have demonstrated advantages in image quality, yet their training cost remains high and the controllability of the generation process is still limited. As a result, targeted environmental modification while maintaining pixel level annotation consistency continues to present a challenge [15,16]. In contrast, multimodal large language models represented by ChatGPT and Gemini have shown strong semantic understanding and instruction following capabilities in image editing tasks [17–20]. Through structured prompting mechanisms, these models can perform targeted modification of environmental appearance while preserving the original geometric structure [19,20], thereby offering a new paradigm for expanding environmental diversity at zero annotation cost. However, the application of model driven environmental synthesis methods of this kind to bridge damage detection remains at an exploratory stage, and systematic experimental evaluation and method comparison are still lacking.

Based on this background, this study proposes a multi weather data generation and evaluation framework for bridge surface concrete spalling detection. To enable a systematic comparison of data construction strategies, three datasets were established, namely, the Base Dataset, the Traditional Augmentation Dataset, and the Generative Artificial Intelligence Dataset. Under a unified instance segmentation model and consistent experimental settings, comparative training and cross dataset testing were performed to evaluate how different data construction strategies affect segmentation performance and generalization ability under complex weather conditions. The results verify the effectiveness of the proposed framework in expanding environmental diversity without additional annotation cost and in improving the stability of automated bridge surface damage detection.

2 Material and Methods

2.1 Overview

To analyze the impact of weather conditions on UAV-based bridge defect detection performance, three datasets were constructed using the same set of bridge surface concrete spalling images acquired from UAV inspections. The Base Dataset was constructed without introducing artificial weather variations. The TA and GAI datasets incorporate weather conditions using different modeling strategies. The three datasets differed only in the adopted weather modeling approach, whereas all other image content and pixel-level annotations remained identical. Each dataset was consistently divided into training, validation, and test sets at a ratio of 6:2:2.

The instance segmentation model YOLOv11 seg was employed as a unified analytical model and trained separately on the Base, TA, and GAI datasets. The inference and evaluation were conducted on the Base and GAI test sets. This cross-testing configuration enables a direct detection performance comparison under conditions without weather interference and complex weather scenarios, thereby facilitating the analysis of the effects of weather variations and different weather modeling strategies on defect detection outcomes. The overall experimental workflow is illustrated in Fig. 1.

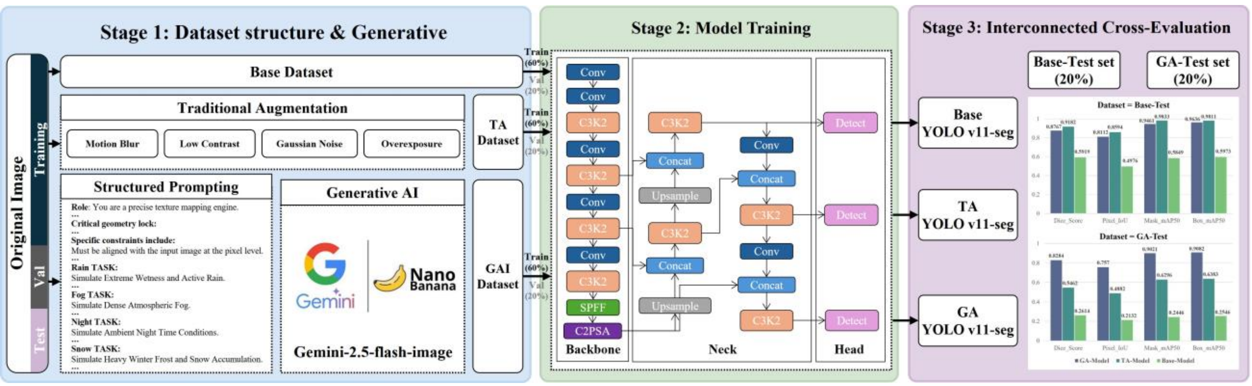

Figure 1. Overview of the experimental framework.

2.2 Data collection

This study used a publicly available bridge façade orthophoto dataset released by the National Information Society Agency of Korea through the AIHUB platform [21]. The dataset covers 27 bridges with different structural types, including PSC I Girder, Steel Box Girder, and Preflex Bridge, and was collected using a DJI Mavic 3 UAV in a close-range inspection mode with a ground sampling distance of 0.3 mm/px and an imaging distance of approximately 1.33 m.

All images were uniformly cropped to 512 × 512 pixels, corresponding to a physical coverage of 15.4 cm × 15.4 cm. Based on these images, a Base Dataset containing only concrete spalling defects was constructed, comprising 551 images and 930 instances annotated with LabelMe polygon masks. Instance areas showed a long-tailed distribution, with small, medium, and large scales accounting for 41.7%, 39.6%, and 18.7%, respectively, where the corresponding pixel ranges were <32 px, 32 to 96 px, and >96 px. The physical size of defects ranged from 0.34 to 79.1 mm. On average, defect regions occupied only 2.85% of the image area, corresponding to a foreground to background ratio of approximately 1:34. The mean bounding box size was 4.1 cm × 1.3 cm, which remained smaller than the image coverage range and therefore preserved sufficient contextual information.

Figure 2. Example of the Base Dataset.

2.3 Traditional Data Augmentation Methods

To simulate weather-related effects on the UAV-based bridge-inspection imagery, a TA was constructed from the Base Dataset. Each original image was augmented without introducing geometric deformations while maintaining full annotation consistency. This process results in 2204 augmented images that were strictly aligned with their original annotations.

The TA Dataset models weather-induced image quality changes through global statistical transformations applied to bridge façade images, including motion blur, low contrast, high brightness or overexposure, and Gaussian noise [9]. Motion blur represents image quality loss caused by UAV motion or wind, with kernel sizes randomly selected between 5 and 15 pixels. Low contrast simulates visibility reduction under hazy or rainy conditions, with scaling factors sampled from 0.4–0.8. High brightness or overexposure model brightness saturation due to strong illumination or surface reflection, with gains randomly adjusted between 1.5 and 2.5. Gaussian noise accounts for sensor noise in low-light or complex environments, with a mean of zero and variance between 20 and 40.

These augmentations introduce environmental variability while preserving overall visual plausibility. However, because transformations are applied globally, they have a limited ability to capture localized appearance changes and their interaction with structural details under complex weather conditions [9,11,12]. Representative examples are shown in Fig. 3.

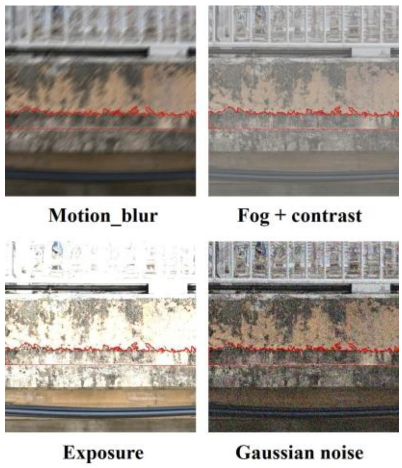

Figure 3. Examples of traditional data augmentation.

2.4 Generative Artificial Intelligence Methods

To compare the capability of generative artificial intelligence and conventional augmentation methods in modeling complex weather conditions, this study adopted a semantic driven image generation approach to simulate rain, haze, snow, and low light nighttime conditions in bridge UAV inspection images, while preserving structural consistency [17–20]. Based on the Base Dataset, one weather generation operation was applied to each original image to construct the generative augmentation dataset, namely the GAI Dataset, resulting in 2,204 images in total.

Image generation was performed using the Gemini 2.5 Flash Image model [19], with sampling parameters set to temperature = 1.0, top p = 0.95, and top k = 64 to maintain visual diversity while limiting geometric distortion. Each image was generated only once and retained directly, without repeated sampling or manual selection, in order to avoid selection bias. Structured prompts were used to introduce only weather related changes in texture, illumination, and visibility, without altering the spatial layout of bridge components or defect regions. To prevent geometric deviation, all generation tasks were constrained by a unified role definition and geometry locking rules. Defect boundaries were not allowed to shift, scale, rotate, or change shape, and only transparent overlay based environmental effects were permitted. The generation rules for different weather types were specified separately through structured prompts [17], and the detailed prompts are listed in Table 1.

To eliminate the risk of data leakage, the 6:2:2 dataset split was performed at the original image level, and all variants derived from the same source image, including the original, TA, and GAI versions, were always assigned to the same subset. Pairwise validation of all 2,204 original to augmented image pairs showed that both mask IoU and Dice coefficient remained exactly 1.000 across all four weather conditions, while the boundary Hausdorff distance was 0.00 pixels. The SSIM ranged from 0.384 to 0.673, indicating that the augmentation substantially changed visual appearance without introducing any annotation misalignment.

Table 1. Structured prompt for weather synthesis.

Role: You are a precise texture mapping engine. Your sole responsibility is to apply environmental texture overlays to the input image. **Critical geometry lock:** No displacement, no deformation, overlay only. **Specific constraints include:** The output image must be aligned with the input image at the pixel level. No cracks or defect boundaries may be moved, scaled, rotated, or reshaped. All effects must be applied exclusively as transparent overlay layers on the existing geometry. The generation style follows a high-resolution industrial inspection imaging standard, preserving the characteristics of raw sensor data.

2.4.1 Rainy Condition Modeling

Rainy-scene synthesis focuses on reproducing the characteristic surface appearance of concrete under wet conditions, as summarized in Table 2. The generation process introduces an overall reduction in brightness, specular reflections caused by a continuous surface water film, and rain streak textures aligned with gravity's direction, while strictly preserving the original geometric structure and defect boundaries.

Table 2. Structured prompt for rain scene synthesis.

TASK: Simulate Extreme Wetness and Active Rain. The concrete surface is completely soaked and appears significantly darker. The surface is covered with a glossy water layer with clear specular reflections. Mirror like reflections of the overcast sky are visible on the wet surface. Add visible vertical rain streaks falling across the image.

As shown in Fig. 4, the generated results successfully reproduce the lighting variations and reduced visibility commonly observed in UAV inspection imagery under real rainfall conditions while preserving the original spatial location and morphology of the defect regions.

Figure 4. Example of rain scene synthesis.

2.4.2 Foggy Condition Modeling

Foggy conditions were modeled by simulating the visual effects of atmospheric scattering, which primarily manifests as reduced contrast and degraded detail visibility. As presented in Table 3, a uniformly distributed cool-gray haze is overlaid during generation to reproduce typical fog-induced appearance degradation without altering the geometric configuration of the defect regions.

Table 3. Structured prompt for fog scene synthesis.

TASK: Simulate Dense Atmospheric Fog. Strong atmospheric scattering significantly reduces visibility. Overall contrast is reduced and shadows are lifted toward gray. A cool gray haze uniformly covers the entire image.

As shown in Fig. 5, the generated images exhibit the characteristic low-visibility imaging features of foggy environments while maintaining the spatial consistency of the defect regions.

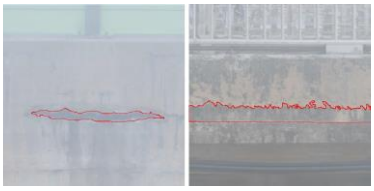

Figure 5. Example of fog scene synthesis.

2.4.3 Snowy Condition Modeling

Snow-scene synthesis aims to capture the coverage effects of frost and snow accumulation on concrete surfaces, as outlined in Table 4. The generation process introduces cold-toned lighting and surface-coverage characteristics while maintaining the original spatial location and geometric shape of the defects, ensuring that only the environmental appearance is modified.

Table 4. Structured prompt for snow scene synthesis.

TASK: Simulate Heavy Winter Frost and Snow Accumulation. A thick layer of frost and snow particles covers the concrete surface. The lighting exhibits a cold bluish winter tone. Snow texture is applied directly over existing cracks and defects.

As shown in Fig. 6, the generated images reflect typical winter environmental characteristics while retaining the spatial continuity of the concrete spalling regions.

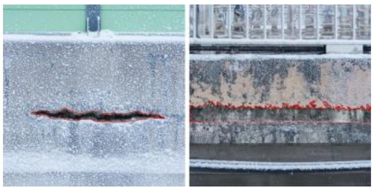

Figure 6. Example of snow scene synthesis.

2.4.4 Nighttime Condition Modeling

Nighttime inspection scenes were generated by modeling appearance degradation under weak-ambient-lighting conditions. As presented in Table 5, the synthesis process reproduces a globally darkened appearance accompanied by increased noise and limited color information without introducing additional light sources or geometric changes, thereby reflecting the typical visual characteristics of nighttime UAV inspections.

Table 5. Structured prompt for night scene synthesis.

TASK: Simulate Ambient Night Time Conditions. The scene is illuminated only by very weak ambient light. No flashlights or direct light sources are allowed. The image is underexposed with strong sensor noise and limited color information.

As shown in Fig. 7, the generated results reflect typical imaging characteristics under low-light nighttime conditions, while maintaining consistency in the defect geometry.

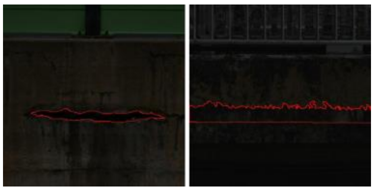

Figure 7. Example of night scene synthesis.

2.5 Model selection

This study formulates the task as instance segmentation to achieve fine grained extraction of concrete spalling regions on bridge surfaces. Concrete spalling typically exhibits irregular boundaries, complex shapes, and multiple spatially separated regions within a single image [2,12]. Compared with other vision tasks, instance segmentation enables pixel level delineation of defect boundaries while distinguishing individual defect instances, making it suitable for morphological and spatial distribution analysis in engineering inspection [8].

YOLOv11n seg was adopted as the instance segmentation model in this study [22]. Within a single stage framework, the model integrates a segmentation branch and generates instance level masks through multi scale feature fusion, allowing object localization and pixel level mask prediction in a unified inference process, as illustrated in Stage 2 of Fig. 1. The model contains 2.84 million parameters, requires 9.6 GFLOPs, and has a file size of only 5.71 MB. Under a 512 × 512 input, its inference latency is approximately 1.0 to 2.5 ms, providing a practical balance between segmentation accuracy and the real time constraints of UAV edge deployment. Based on these characteristics, YOLOv11n seg was consistently used as the segmentation model in all data augmentation comparison experiments.

2.6 Experimental Strategy

All experiments were conducted under the same computational setting, with the data construction strategy as the only variable. The platform consisted of Windows 11 Pro, Python 3.9, and PyTorch 2.8.0, running on an NVIDIA GeForce RTX 5080 GPU and an AMD Ryzen 9 9950X processor. YOLOv11 seg initialized with the pretrained weight yolo11n-seg was used in all groups.

During training, all images were resized to 512 × 512 pixels. The model was trained for up to 200 epochs using AdamW, with a learning rate of 0.001, weight decay of 0.0005, batch size of 16, and random seed of 42. Early stopping was applied when validation performance did not improve for 20 consecutive epochs. No additional random augmentation was introduced.

Evaluation was performed on the same test sets using Box mAP50, Mask mAP50, pixel level IoU, and Dice coefficient [8]. A prediction was regarded as correct when its IoU with the ground truth was not less than 0.50. For a predicted region *P* and and ground truth region *G*,

$$IoU(P,G) = \left|\frac{P \cap G}{P \cup G}\right| \tag{1}$$

Pixel level consistency was measured on the predicted mask M_{pred} and ground truth maskM_{gt}:

$$IoU_{pixel} = \left|\frac{M_{pred} \cap M_{gt}}{M_{pred} \cup M_{gt} + \varepsilon}\right| \tag{2}$$

$$Dice = \frac{2\left|M_{pred} \cap M_{gt}\right|}{\left|M_{pred}\right| + \left|M_{gt}\right| + \varepsilon} \tag{3}$$

where ε denotes a numerical stabilization term. Final pixel level results were obtained by averaging all test samples.

3 Result

3.1 Cross-Dataset Segmentation Performance

All models exhibited stable convergence without obvious overfitting or underfitting, their quantitative performance on the test set is presented in Fig.8.

On the GA test set, the GA model achieved the best segmentation performance, with a mask mAP50 of 0.902 and a Dice coefficient of 0.828. These values are substantially higher than those of the TA model (0.630 and 0.546, respectively), while also being markedly superior to those of the BASE model (0.245 and 0.261, respectively). These results indicate that models trained with generative augmentation data are more effective at learning damage representations in complex environments when the test distribution differs significantly from the training distribution.

On the BASE test set, the TA model attained the highest segmentation accuracy, achieving a mask mAP50 of 0.983 and a Dice coefficient of 0.918. Meanwhile, the GA model maintained comparable performance on the same test set, with a mask mAP50 of 0.946 and a Dice coefficient of 0.877, without exhibiting noticeable performance degradation. This suggests that the introduction of generative augmentation data does not compromise the segmentation capability of the original data distribution.

Performance differences across the two test sets further demonstrate that the GA model maintains stable segmentation accuracy under conditions in which the training and testing distributions are inconsistent, whereas the TA model is more sensitive to test data distribution variations. These experimental results validate the effectiveness of generative-AI-augmented data in enhancing the cross-scenario generalization capability of instance segmentation models.

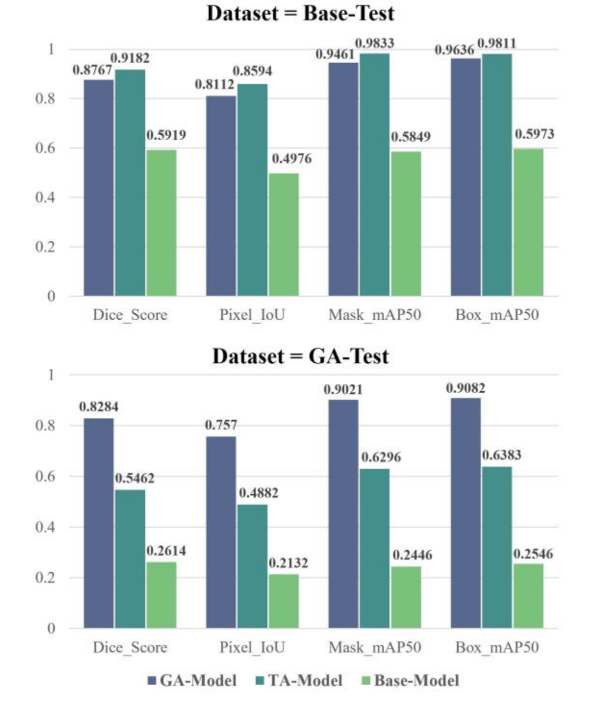

Figure 8. Cross-dataset segmentation performance.

3.2 Qualitative Comparison under Diverse Environmental Conditions

To provide a visual comparison of instance segmentation behavior under different training strategies, Fig. 9 presents representative segmentation results and their corresponding Dice scores under baseline, rain, night, snow, and haze conditions. Under the baseline test condition, the Dice scores of the Base, TA, and GA models were 0.90, 0.94, and 0.9423, respectively, indicating only minor performance differences.

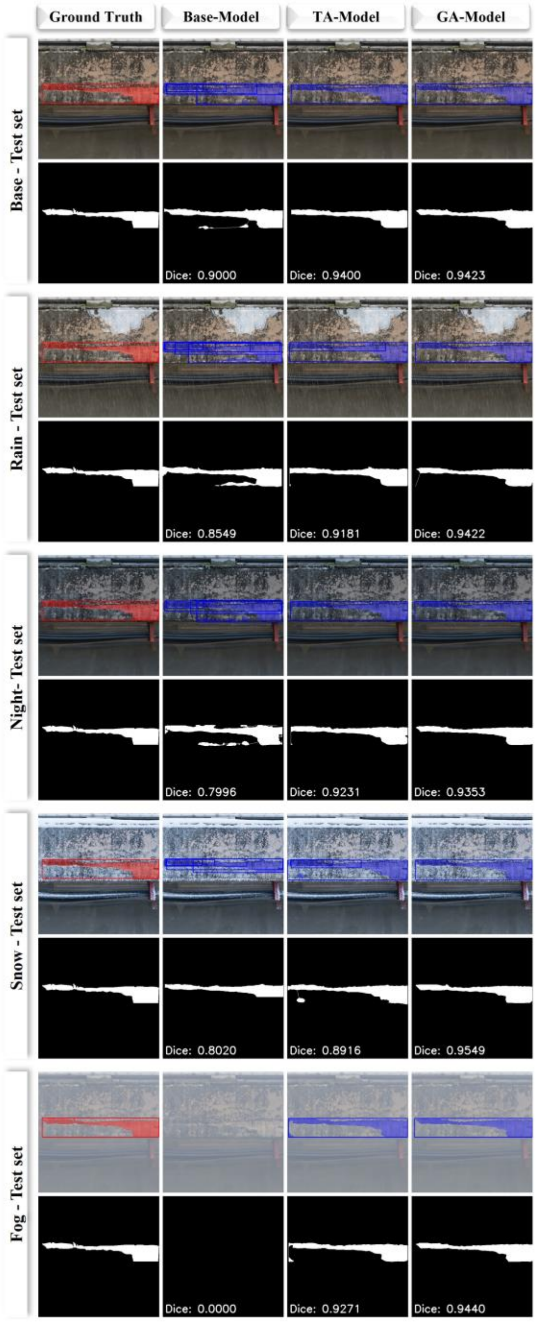

Figure 9. Segmentation examples under different weather.

As environmental conditions became more challenging, the performance gap between models became increasingly evident. This divergence is fundamentally associated with the different distribution modeling mechanisms of the two augmentation strategies. Traditional augmentation methods, such as motion blur, contrast reduction, and Gaussian noise, are essentially spatially uniform global statistical transformations. As a result, they cannot reproduce the localized appearance changes that are semantically coupled with scene content in real weather conditions, such as wet specular reflection on concrete surfaces in rain or uneven occlusion of defect regions under snow. In rain and night scenes, the Dice scores of the Base model decreased to 0.8549 and 0.7996, whereas both the TA and GA models remained above 0.91. Under snow conditions, the Base model achieved a Dice score of only 0.8020, the TA model reached 0.8916, and the GA model further improved to 0.9549, showing superior regional continuity and boundary consistency.

The difference was most pronounced under haze conditions. Real atmospheric scattering exhibits depth dependent spatial characteristics, where distant details are attenuated more severely than near field regions, producing a nonuniform visibility gradient. Traditional augmentation applies only uniform contrast degradation and therefore cannot reproduce this physical property. Consequently, the Base model completely failed to identify the defect region, with a Dice score of 0.00. The TA model detected only the main damage area and showed insufficient boundary completeness, with a Dice score of 0.9271. In contrast, the GA model maintained stable performance by capturing semantically consistent scattering patterns, achieving a Dice score of 0.9440.

These Dice trends are consistent with the quantitative cross dataset evaluation results, indicating that the main advantage of generative augmentation lies in modeling the semantic coupling between weather conditions and scene content, rather than merely expanding pixel level statistical variation.

4 Conclusion

In this study, conducted within the context of UAV-based bridge inspection, systematically evaluated the impact of different weather condition modelling strategies on the instance segmentation performance of concrete spalling. Under a unified model architecture and consistent training configuration, a comparative analysis was performed using real inspection data, traditionally augmented data, and multi-weather datasets constructed via generative AI. The experimental results demonstrated that under complex environmental conditions, including rain, nighttime, haze, and snow, the model trained with generative augmentation data exhibited the most stable performance in terms of segmentation accuracy and pixel-level consistency. In the complex environment test set, this model achieved a mask mAP50 of 0.902 and a Dice coefficient of 0.828, while reaching 0.946 and 0.877 on the original test set, respectively.

Further analysis shows that the generative AI based multi weather dataset reduced training bias toward a single environment by combining semantic driven synthesis with geometry preserving constraints, thereby covering inspection conditions that are difficult to collect systematically while maintaining defect location and annotation consistency. The contributions of this study are threefold. First, it proposes a zero-annotation cost strategy for multi weather synthesis, in which structured prompts and geometry locking enable environmental appearance modification without changing pixel level labels. Second, it establishes a controlled three group experimental framework with a shared model architecture and training configuration, allowing the independent effect of data construction strategy on segmentation performance to be quantitatively isolated. Third, the model trained with generative augmentation achieved a Mask mAP50 of 0.902 on the synthetic weather test set, representing a 43.2% improvement over traditional augmentation, while maintaining 0.946 on the original test set. This result confirms that environmental diversity can improve model performance without sacrificing baseline accuracy. Because the framework requires no additional annotation effort and can be integrated into existing inspection workflows, and because YOLOv11 seg supports efficient single stage inference, the proposed approach also shows potential for edge deployment.

This study still has several limitations. First, the annotation reuse strategy implicitly assumes that weather synthesis changes appearance only and does not alter defect visibility. This assumption may not hold in extreme conditions, such as thick snow physically covering the defect, dense haze, or very low illumination that makes the defect visually indistinguishable. In such cases, direct reuse of the original labels may introduce label noise. Second, the experiments focus only on concrete spalling and do not verify applicability to multi defect detection tasks. In addition, the generative dataset was developed mainly for typical weather conditions and does not yet cover continuous environmental transitions that may occur in real inspections. Third, only a single instance segmentation model was evaluated, and the adaptability of different architectures to generative augmentation remains unclear. Future work should incorporate visibility aware label adjustment to handle severely degraded conditions, extend the framework to multi defect and multi component inspection tasks, and combine finer grained environmental modeling with multi model comparisons to further define the practical scope of generative data augmentation in intelligent bridge inspection.

Acknowledgement

This research was funded by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (Ministry of Science and ICT) (grant number RS202400356697).

Reference

[1] C. Huang, Y. Zhou, and X. Xie. Intelligent diagnosis of concrete defects based on improved Mask R CNN. *Applied Sciences*, 14:4148, 2024.

[2] P. Huethwohl, R. Lu, and I. Brilakis. Multi classifier for reinforced concrete bridge defects. *Automation in Construction*, 105:102824, 2019.

[3] H. Zoubir, M. Rguig, M. E. Aroussi, A. Chehri, R. Saadane, and G. Jeon. Concrete bridge defects identification and localization based on classification deep convolutional neural networks and transfer learning. *Remote Sensing*, 14, 2022.

[4] Federal Highway Administration. National Bridge Inventory database. On-line: https://www.fhwa.dot.gov/bridge/nbi.cfm, Accessed: 01 04 2025.

[5] S. Dorafshan and M. Maguire. Bridge inspection human performance, unmanned aerial systems and automation. *Journal of Civil Structural Health Monitoring*, 8:443 to 476, 2018.

[6] R. Li and Z. Wang. AGSAM Net UAV route planning and visual guidance model for bridge surface defect detection. *Image and Vision Computing*, 154:105416, 2025.

[7] F. Wang, Y. Zou, X. Chen, C. Zhang, L. Hou, E. del Rey Castillo, and J. B. P. Lim. Rapid in flight image quality check for UAV enabled bridge inspection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 212:230 to 250, 2024.

[8] C. Y. Liu and J. S. Chou. Bayesian optimized deep learning model to segment deterioration patterns underneath bridge decks photographed by unmanned aerial vehicle. *Automation in Construction*, 146:104666, 2023.

[9] T. Askarzadeh and R. Bridgelall. Cost efficiency and effectiveness of drone applications in bridge condition monitoring. *Infrastructures*, 10, 2025.

[10] Y. Yang, S. Yang, C. Li, Y. Wang, X. Pi, Y. Lu, and R. Wu. Insulator defect detection under extreme weather based on synthetic weather algorithm and improved YOLOv7. *High Voltage*, 10:69 to 77, 2025.

[11] S. Wang. Effectiveness of traditional augmentation methods for rebar counting using UAV imagery with Faster R CNN and YOLOv10 based transformer architectures. *Scientific Reports*, 15:33702, 2025.

[12] D. Amirkhani, M. S. Allili, L. Hebbache, N. Hammouche, and J. F. Lapointe. Visual concrete bridge defect classification and detection using deep learning a systematic review. *IEEE Transactions on Intelligent Transportation Systems*, 25:10483 to 10505, 2024.

[13] Z. Liang, M. Zhang, X. Zhao, W. Liu, and Y. Wang. Robust traffic sign detection in adverse weather via hybrid vision transformer and GAN based augmentation. *In Proceedings of the 7th International Conference on Data Driven Optimization of Complex Systems*, Taiyuan, China, pages 340 to 344, 2025.

[14] B. Li, H. Guo, and Z. Wang. Data augmentation using CycleGAN based methods for automatic bridge crack detection. *Structures*, 62:106321, 2024.

[15] W. Shen, D. Zeng, Y. Zhang, X. Tian, and Z. Li. Image augmentation for nondestructive testing in engineering structures based on denoising diffusion probabilistic model. *Journal of Building Engineering*, 89:109299, 2024.

[16] S. Yu, J. Li, H. Zheng, and H. Ding. Research on predicting building facade deterioration in winter cities using diffusion model. *Journal of Building Engineering*, 111:113365, 2025.

[17] Y. Lee, G. Kang, J. Kim, S. Yoon, and J. Jeon. Generative AI driven data augmentation for enhanced construction hazard detection. *Automation in Construction*, 177:106317, 2025.

[18] Y. Qian, E. Bocek Rivele, L. Song, J. Tong, Y. Yang, J. Lu, W. Hu, and Z. Gan. Pico Banana 400K a large scale dataset for text guided image editing. *arXiv preprint*, arXiv 2510.19808.

[19] G. Comanici, E. Bieber, M. Schaekermann, I. Pasupat, N. Sachdeva, I. Dhillon, M. Blistein, O. Ram, D. Zhang, E. Rosen, and others. Gemini 2.5 pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint*, arXiv2507.06261, 2025.

[20] J. Betker, G. Goh, L. Jing, T. Brooks, J. Wang, L. Li, L. Ouyang, J. Zhuang, J. Lee, Y. Guo, and others. Improving image generation with better captions. *Computer Science*, 2:8, 2023.

[21] National Information Society Agency. Bridge exterior inspection rectified facade image dataset. On-line: https://www.aihub.or.kr, Accessed: 01 04 2026.

[22] R. Khanam and M. Hussain. YOLOv11 an overview of the key architectural enhancements. *arXiv preprint*, arXiv:2410.17725, 2024.